PARTHENOS

Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

A Webinar of the PARTHENOS eHumanities and eHeritage Series

#lovedata2018

How to work successfully with eHumanities and eHeritage Research Infrastructures: The Devil is in the Details!

(Klaus Illmayer and Marie Puren)

Webinar start: **13.02.2018, 11:00 CET**

Moderation: Ulrike Wuttke
(University of Applied Sciences Potsdam)

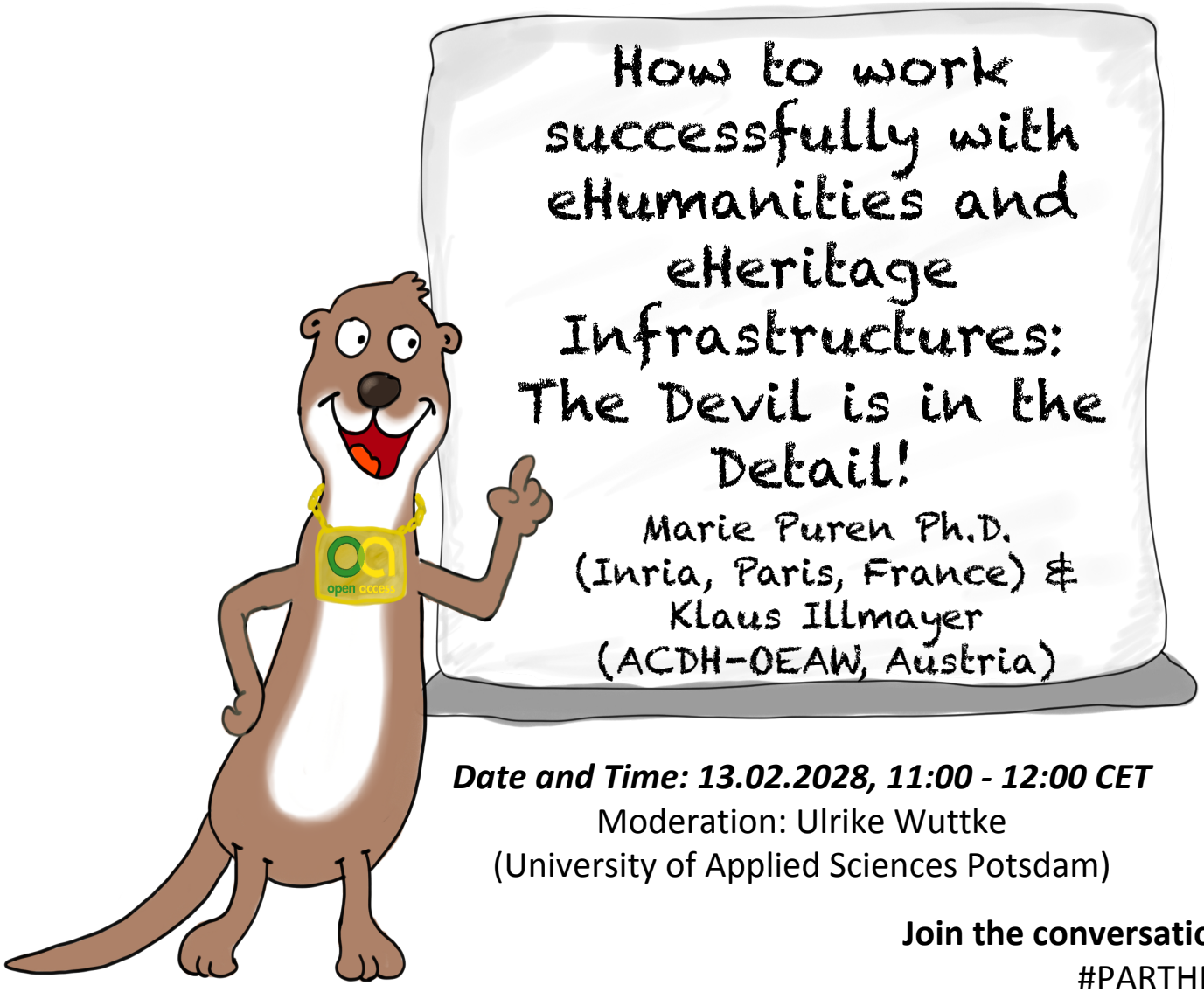
FH;P Fachhochschule Potsdam University of Applied Sciences

Join the conversation on Twitter:
#PARTHENOSWebinar
@Parthenos_EU

PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

FH;P Fachhochschule Potsdam University of Applied Sciences

PARTHENOS eHumanities and eHeritage Webinar Series



Date and Time: 13.02.2028, 11:00 - 12:00 CET

Moderation: Ulrike Wuttke
(University of Applied Sciences Potsdam)

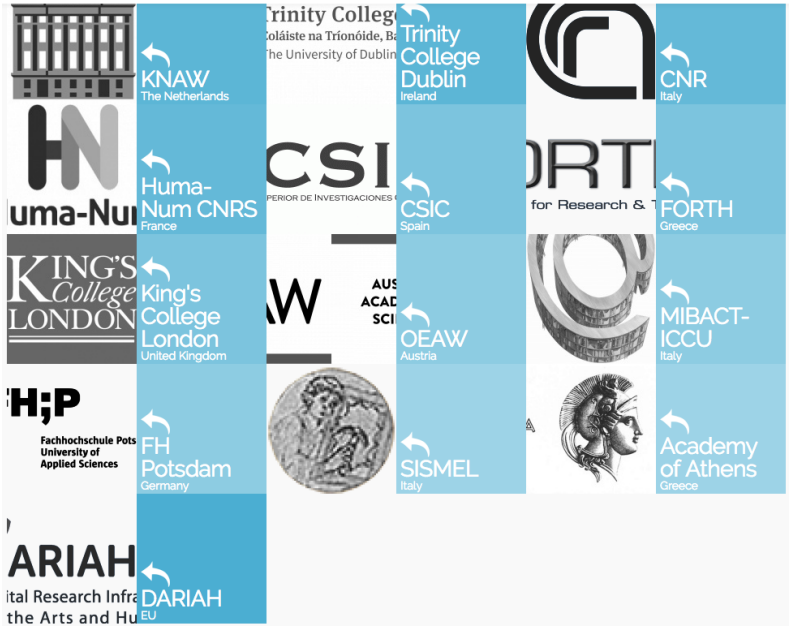
Join the conversation on Twitter:
#PARTHENOSWebinar
@Parthenos_EU

Some Housekeeping…

- As a participant you will be **muted** and not be seen throughout the webinar.
- For **questions** and **remarks** please use the **chat**. For questions, please use the word **question**.
- Your questions are going to be answered by the trainers during / after the presentation.
- If you have **sound problems**, please test your technical settings. You have to click on the **speaker symbol** on top to to be able to hear (it has to be green).
- **Help us improve**: Follow up e-mail with a link to a short **feedback survey**.

Some words about PARTHENOS…

- **PARTHENOS** stands for: Pooling Activities, Resources and Tools for Heritage eResearch Networking, Optimization and Synergies
- PARTHENOS is a **Horizon 2020** project with the aim to **strengthen the cohesion of Heritage related E-research**
- Running time: 1 May 2015 - 30 April 2019
- PARTHENOS has 16 partners from 9 European countries
- PARTHENOS Coordinator: PIN Scrl (Italy)
- The PARTHENOS webinar series is a **cross PARTHENOS training effort**



PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

FH;P Fachhochschule Potsdam University of Applied Sciences

How to work successfully with eHumanities and eHeritage Infrastructures: The Devil is in the Details

Trainers

Dr Marie Puren
(Inria, Paris, France)

Klaus Illmayer
(ACDH-OEAW, Austria)

Picture Klaus Illmayer:
CC-BY 4.0 Sandra Lehecka

Moderation

FH;P Fachhochschule Potsdam University of Applied Sciences

Dr Ulrike Wuttke
(University of Applied Sciences Potsdam, Germany)

Warm up: Tell us where you are

Klaus Illmayer (Vienna - Austria, ACDH-OEAW)
Marie Puren (Paris - France, INRIA)
Ulrike Wuttke (Potsdam - Germany, FHP)

Now tell us the **place** from where you are participating:

- Please use the **chat function** (you can find it at the bottom on the right side)
- If you don't see it, please **close the full screen**
- You can use the chat function to **ask questions**. We will answer it after our presentation in the Q&A section

Covered topics

- e-Infrastructures
- Preparing data
  - FAIR data principles
  - Standards (Standardization Survival Kit - SSK)
- Conclusions
- Q&A

eHumanities and eHeritage Research Infrastructures

Definitions - let's have a look at PARTHENOS:

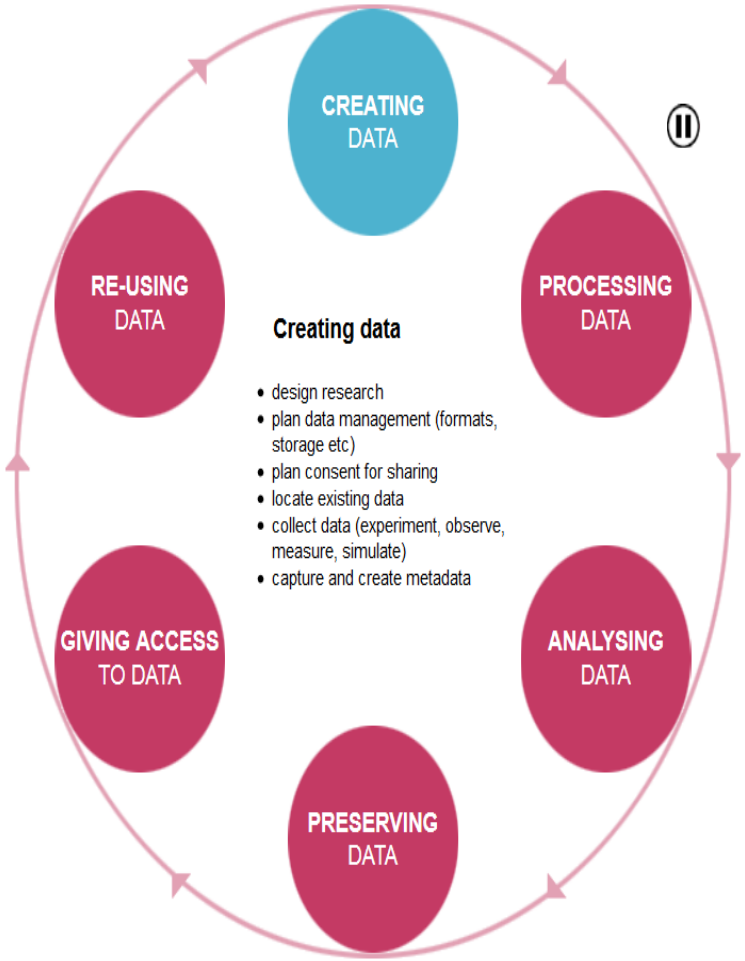
- "Pooling **Activities, Resources and Tools** for Heritage E-research Networking, Optimization and Synergies"
- Focus on **research communities** from humanities and heritage
- **Connect** analog and digital research, **connect** digital infrastructures; **harmonize** standards; **share** best practices
- Examples: Joint Resource Registry, Standardization Survival Kit (SSK), collection of recommendations and best practices, …
- PARTHENOS itself is an **e-Infrastructure**

What is a Research e-Infrastructure?

- A **community** driven **digital environment** for research
  - Allowing community members to work together in means of **shared** communication, resources, tools and services
  - More extended: Virtual Research Environments (VRE)
    - Bundle communication, resources, tools and services to establish a **digital workflow**, tailored to the needs and practices of the community
  - Even more extended: Combine different communities
- Ideally: A technical framework to support the **research data life cycle**

Research data life cycle #1

UK Data Archive Research Data Lifecycle, http://www.data-archive.ac.uk/create-manage/life-cycle



Research data life cycle #2

- Steps of the life cycle rely on **different tools and services**
- **Different practices** behind these steps, e.g. preserving data with a digital archive like ACDH's ARCHE
- **Data** that goes through the life cycle may **initiate a new life cycle**
- **Combination** of different tools, services and e-Infrastructures (and you may need to rearrange it)

Scope of e-Infrastructures

- Enabling **sharing**
- Scales up: **From** a simple communication **tool to** a **platform** that deliver services, integrate tools and supports work with data
- Different **sizes** of e-Infrastructures: smaller ones may fit better for specific research purposes
- A small e-Infrastructure: your local computer and the environment you set up for doing your research work

Examples of e-Infrastructures

- Research projects using digital tools and services
- Institutions: ACDH-OEAW, INRIA, ...
- National consortia: DARIAH-DE, CLARIAH-NL, DHA, ...
- Transnational networks: CLARIN ERIC, DARIAH-EU, RDA, OpenAIRE, ...
- Data hubs: Europeana, HAL, Zenodo, GitHub, ...

What is the background of an e-Infrastructure?

- Combination of **hardware and software** that enables shared work
- **Maintenance** of this combination
- Further **development** and **integration** of services
- Creating an **user interface** that supports participation
- The more participants and services, the more **complex** this becomes
  - Combining software for a digital workflow can get tricky
  - Integrating different communities, research methods, and data can get confusing

How an e-Infrastructure should look like #1

Users of an e-Infrastructure should not be bothered with the technical background:

- I**ntegrating** hardware and software **silent**: researchers want to work on their data and not on the setup of the e-infrastructure
- The technical configuration needs to be **transparent**, preventing a “black box”
- Providing an **easy entry point** so that users can work immediately

Requirements for working with an e-Infrastructure

- Have a **research question** and even better, have **research data**
- If you have a **running project** or a **project in preparation**, **connect with an e-infrastructure** on how to integrate your data
  - Understand how to integrate your data: Which standards are in use? Are there access points where you can put or get data from (API)?
- Some e-Infrastructures **support** you in **creating data**, which is of relevance if you don't have data already gathered
- **Define** your **requirements** for data to compare different e-Infrastructures

How an e-Infrastructure should look like #2

- A **rich combination** of data, services and tools that allows to **create data, to gather data and to connect data** for research questions
- Supports the **implementation** of research data life cycle steps
- Valuable interfaces to **integrate** research data
- **Facilitate** research and **harvest** research data
- Be aware: There is **no one-approach-fits-it-all** e-Infrastructure

How to connect with an e-Infrastructure?

- **Ask** colleagues
- **Discuss** in your community
- **Connect** with local support institutions
- **Contact** an e-infrastructure
- **Compare** e-Infrastructures to identify the best fitting approach
- **Combine** e-Infrastructures with other e-Infrastructures, with (self-developed) services, and with tools
- **Create** a digital workflow

Further information on e-Infrastructures

- PARTHENOS Training: Introduction to Research Infrastructures

  http://training.parthenos-project.eu/sample-page/intro-to-ri/
- Open Data Institute: What is data infrastructure?
  https://theodi.org/what-is-data-infrastructure
- European Commission: Policies - e-Infrastructures

  https://ec.europa.eu/digital-single-market/e-infrastructures
- European Commission: Infrastructure publications

Take away messages e-Infrastructures

- E-Infrastructures **connect** you with stakeholders
- For Researchers:
  - e-Infrastructures **support** you
  - … give you **access to data** for research
  - … **create impact** for research
- For Cultural Heritage Institutions:
  - e-Infrastructures **raise disseminiation** of data
  - … enable researchers to **work** with data
  - … deliver new **insights** into data
- To succeed, it requires **data quality** => **community effort**

Data quality is important

- Identify for your project **what is data**
- Create **structured** data
- Make a **data management plan** (DMP)
- Publish your data as **open** as possible

Data quality is about:

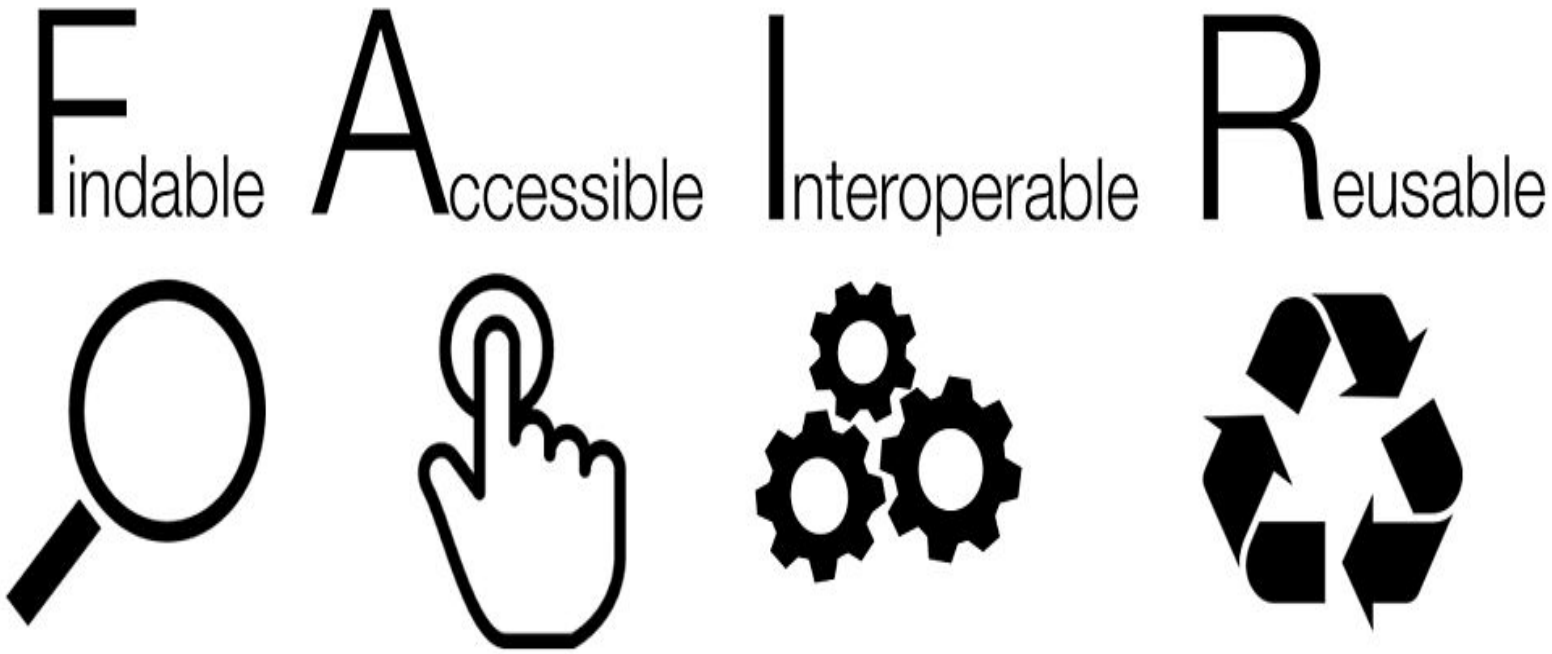
- Apply **FAIR** data principles
- Use **standards**

Key to success: **Preparation** and **continuous** work on data quality

Are you aware of the FAIR data principles?

- We prepared a poll
- Please close the full screen window to see it
- Take your vote!

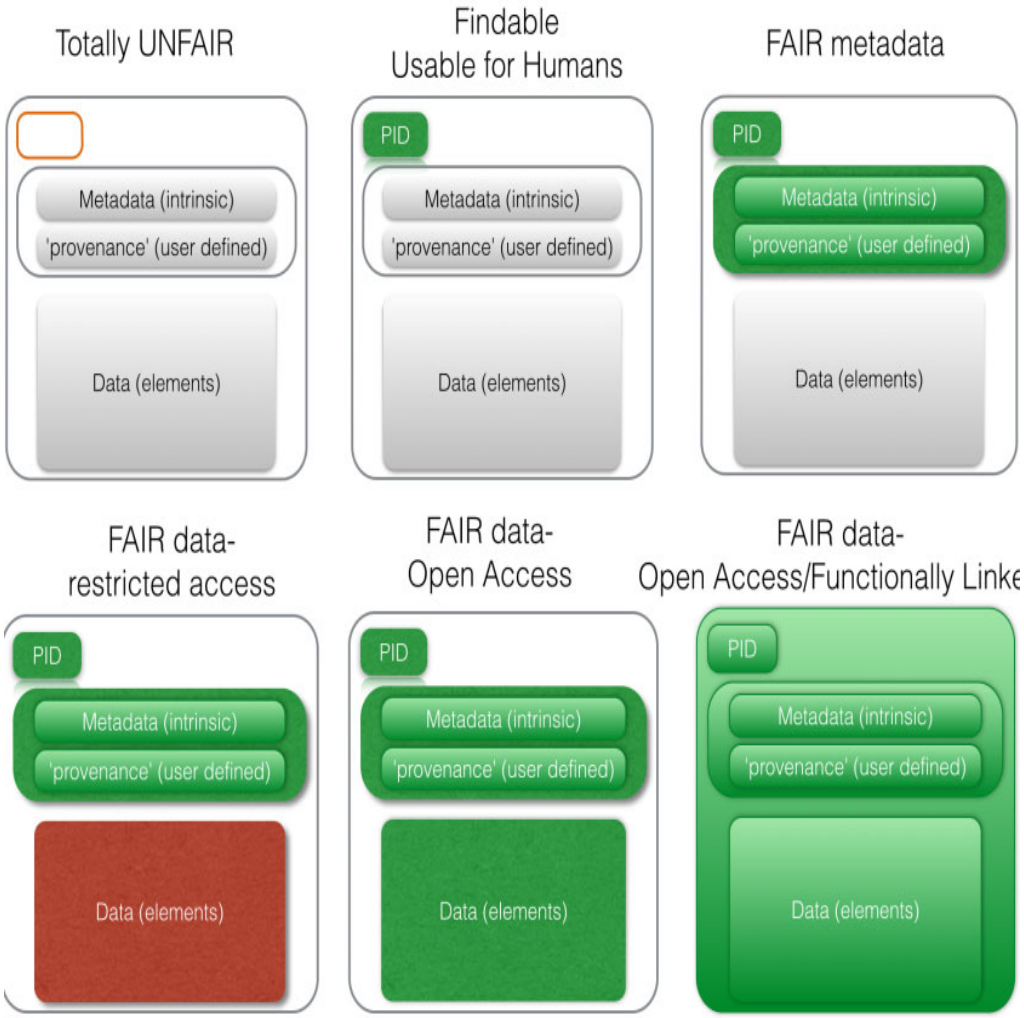
What are the FAIR data principles?



The FAIR data principles

- Compiled by FORCE11
  - https://www.force11.org/group/fairgroup/fairprinciples
  - Guiding principles: https://www.force11.org/fairprinciples
- **Generic** principles to **raise data quality**
- Applying them **supports the research data life cycle**
- … and the **further development of e-Infrastructures**
- Different ways to **implement** them
- Asks for **involvement** of all stakeholders

Data as increasingly FAIR Digital Objects



Taken from FORCE11 website:
https://www.force11.org/fairprinciples

Further information on FAIR data principles

- PARTHENOS training: Manage, Improve and Open up your Research and Data: http://training.parthenos-project.eu/sample-page/manage-improve-and-open-up-your-research-and-data/
- FAIRsharing: https://fairsharing.org/
- GO FAIR: https://www.go-fair.org/
- Outcome PARTHENOS Work Package 3: Support tool for the implemenation of common policies
- Make your data FAIR!

How can I make my data FAIR?

A **Data Management Plan**
"helps 2020 beneficiaries make their research data
Findable,
Accessible,
Interoperable and
Reusable
(**FAIR**) […]."

H2020 Programme Guidelines on FAIR Data Management in Horizon 2020, Version 3.0, 26 July 2016, p.3.

How can I make my data FAIR?

Write a Data Management Plan (DMP)

(Picture Research Lifecycle: PARTHENOS CCO)

What is a Data Management Plan (DMP)?

What is a Data...

"**Research data** is data that is **collected**, **observed**, or **created**, for purposes of analysis to produce **original research results**."

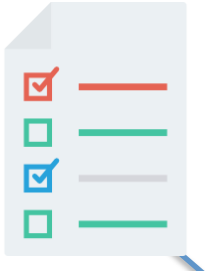
Margaret E. Henderson,

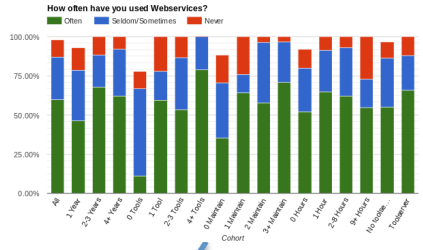
Data Management: A practical guide for librarians, 2016, p.2.

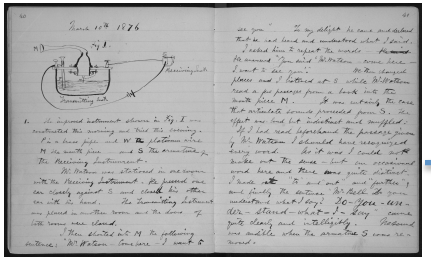
PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

FH;P Fachhochschule Potsdam University of Applied Sciences

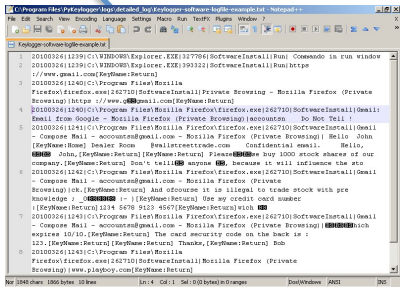
What is a Data...







Research data



…Management…

- An ongoing **maintenance** (backup, migration, conversion…)
- An **action plan** in terms of data quality, technical feasibility and financial viability

- Identifying and making **visible** the actions to be conducted
- Planning **key stages**, **deadlines** and **critical time periods**

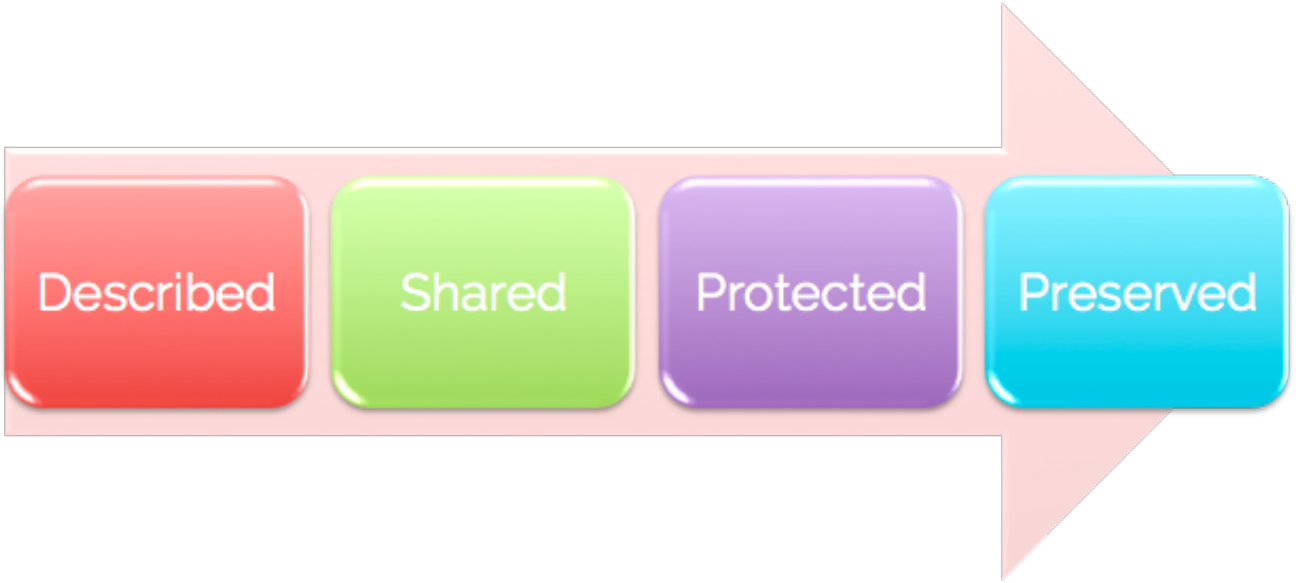
…Management…

Manage your data, don't let your data manage you!

...Plan?

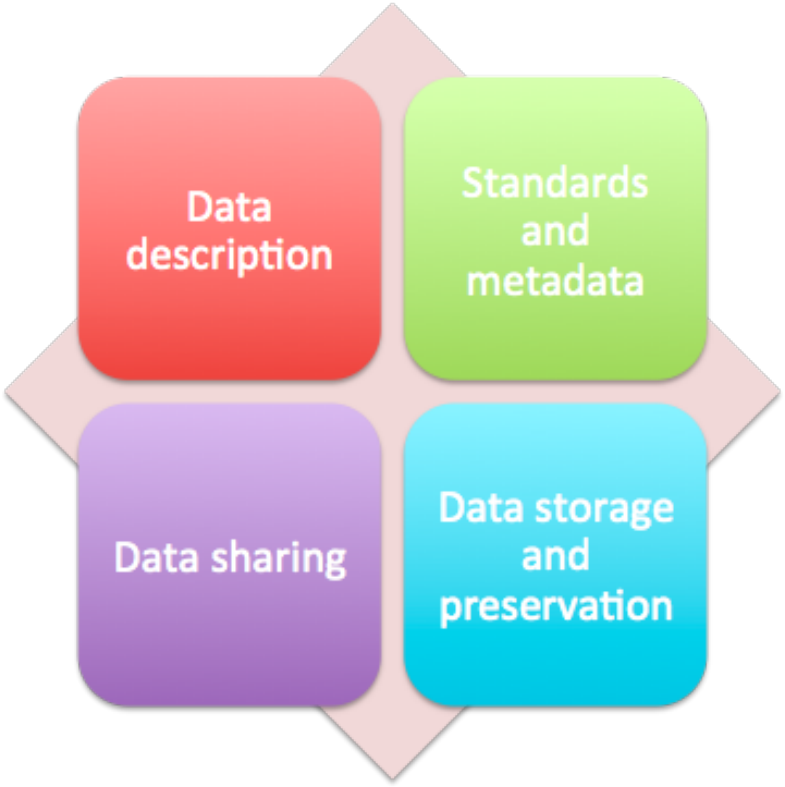
…Plan?

When you create a DMP, you explain how data will be:



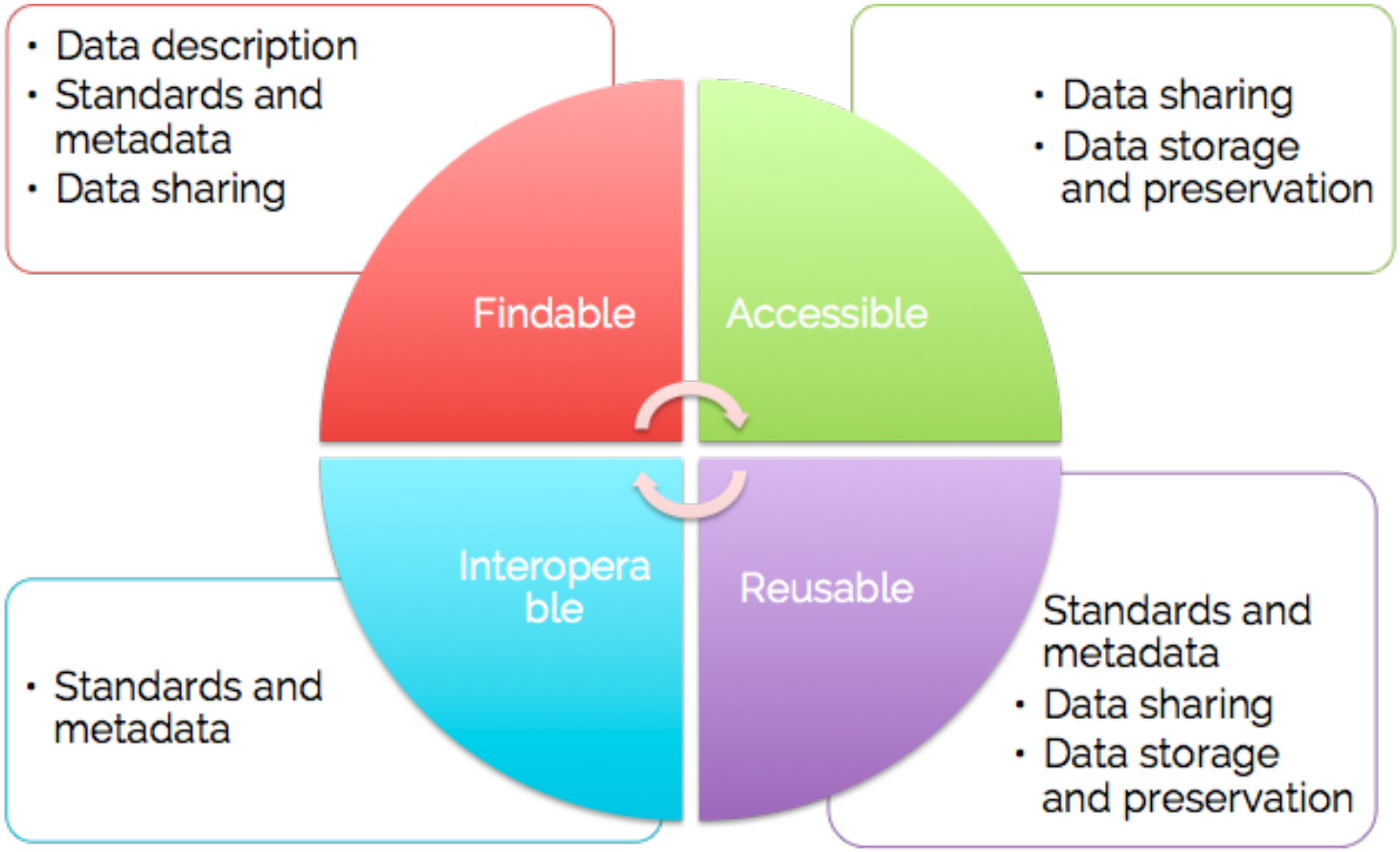
And in practice?

Why would my data be FAIR with a DMP?

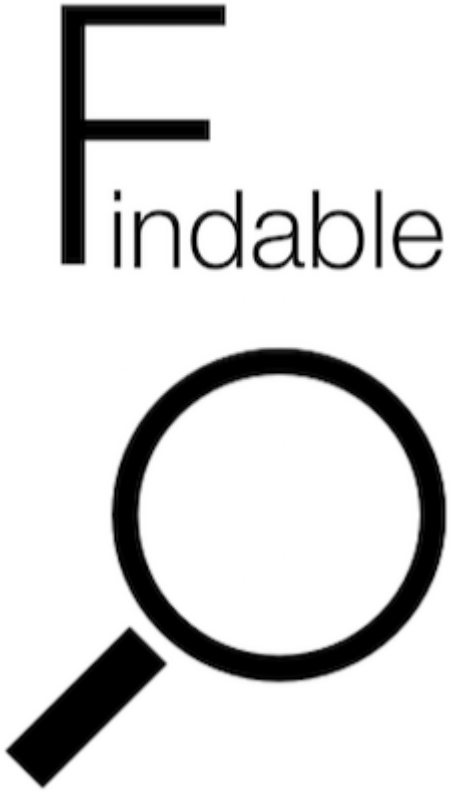


And in practice?

Why would my data be FAIR with a DMP?



In practice, my data are findable because I...



- Pick a **naming convention** (and stick to it)
- Provide clear **version numbers**

EXAMPLE:
title_project_identifier_v2_yyyymmdd

- Create **metadata** describing your data
- Include **appropriate standards** for the content and format of your data

In practice, my data are findable because I...

Create metadata (among other things)

Metadata are data about data

Alice's adventures in Wonderland ; Through the looking glass

Auteur : Lewis Carroll

Éditeur: New York : Barnes & Noble, Inc., 2015.

Édition/format: Livre imprimé : Fiction : Public jeunesse : Anglais Voir toutes les éditions et tous les formats

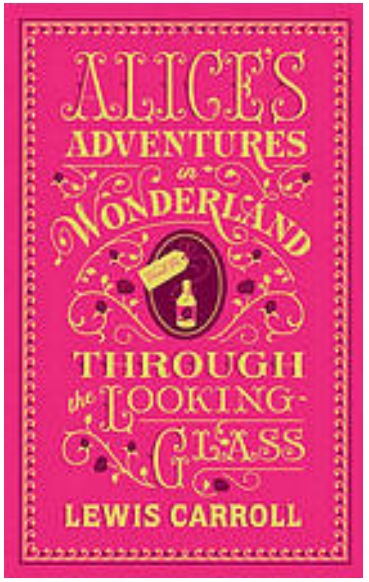
Résumé: Lewis Carroll's novels Alice's Adventures in Wonderland and Through the Looking Glass (first published in 1865 and 1871, respectively) have entertrained readers young and old for more than a century. Their magical worlds, amusing characters, clever dialogue, and playfully logical illogic epitomize the whit and whimsy of Carroll's writing. Alice's Adventures in Wonderland transports you down the rabbit-hole into a Lire la suite...

Évaluation: ☆☆☆☆☆ (pas encore évalué) 0 avec des critiques - Soyez le premier

Sujets Alice -- (Fictitious character from Carroll)
Alice (Fictitious character : Carroll) -- Juvenile fiction.

Plus comme ceci Listes d'utilisateurs Ouvrages semblables

Provides information on



Worldcat

Real world

In practice, my data are accessible because I...

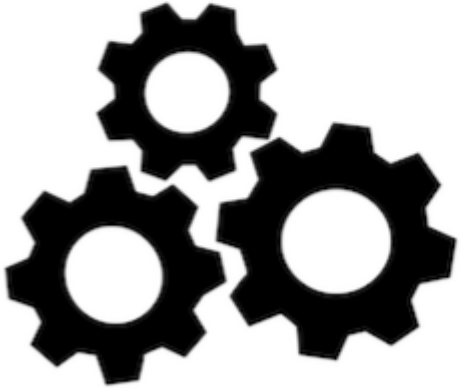
- Deposit my data and associated metadata in a **research data repository** that will identify my data with a unique persistent identidire (such as DOI or Digital Object Identifier) (E.g. Zenodo)

TIP – I can find a research data repository with re3data or "Registry of Researcha data Repositories" that is a catalog of research data repositories.

- Precise which **methods and software tools** are needed to access my data

In practice, my data are interoperable because I...

Interoperable

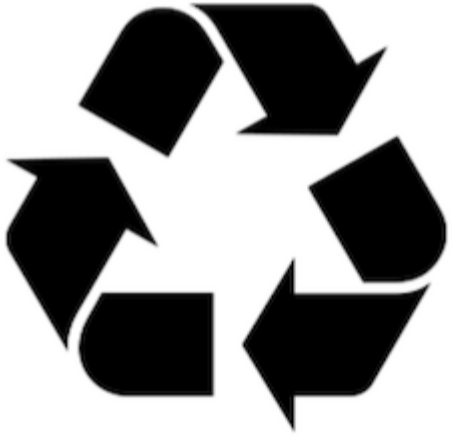


- Use **standards** to modelize and describe my data
- Use **standards vocabularies or ontologies** for the metadata I created

DEFINITION – In addition to metadata and standardised metadata schemas, research infrastructures can also use other forms of "knowledge representation system" to enhance the researcher's experience of the interoperable data they present. When we talk about 'Knowledge Representation Systems' in research infrastructures, we usually mean a specific category of hierarchical systems of terms known more commonly as an 'ontology'. Before the digital age, philosophers referred to an ontology as "the study of the kinds of things that exist". Ontologies are similar to taxonomies […].", PARTHENOS, What Are Knowledge Representation Systems and 'Ontologies'?

In practice, my data are reusable because I...

Reusable



- Have obtained the **informed consent** for people participating to the research (if needed)
- **License my data** with a licence as open as possible

TIP– I can use the Creative Commons framework to combine 4 license types, precising the rights I give up or not.

- **Store data** in a long-term archive.

The Devil is in the details

If you want make your data FAIR!

Standards are crucial for research

"Standards are a **key to great digital research**, which helps to discover and understand our cultural and societal life."

Laurent Romary, chairman of the Technical Committee "Terminology and other language and content resources" of the International Organization for Standardization (ISO)

What are standards?

- They **inform** you about practices, protocols, artefact characteristics or data formats.
- They can be used as **reference** when you work with colleagues from your field, so that you can produce **comparable or interoperable** results.
- They are not **regulation**, you don't have to follow them (but you should).
- They are published and maintained by **standards organizations** such as

What are standards?

To be called "standards", they must fulfill three requirements:

- They express a **consensus**
- They are published and easily **accessible**
- They are **maintained**

Why standards?

Truan, Naomi. 2016. Parliamentary Debates on Europe at the Deutscher Bundestag (1998-2015) [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). https://hdl.handle.net/11403/de-parl

Truan, Naomi. 2016. Parliamentary Debates on Europe at the House of Commons (1998-2015) [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). https://hdl.handle.net/11403/uk-parl

Truan, Naomi. 2016. Parliamentary Debates on Europe at the Assemblée nationale (2002-2012) [Corpus]. ORTOLANG (Open Resources and TOols for LANGuage). https://hdl.handle.net/11403/fr-parl

List of metadata in the <teiHeader>:

- <fileDesc>
- <encodingDesc>
- <profileDesc>
- <revisionDesc>

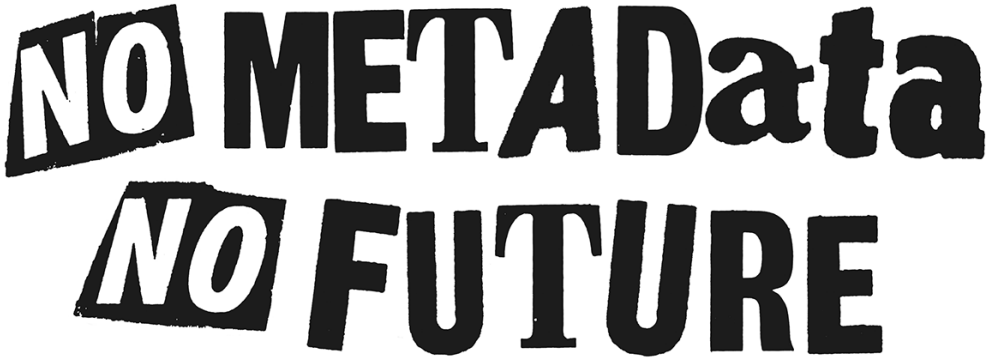
Why standards?

- A metadata schema consists of a definite set of characteristics to describe data.

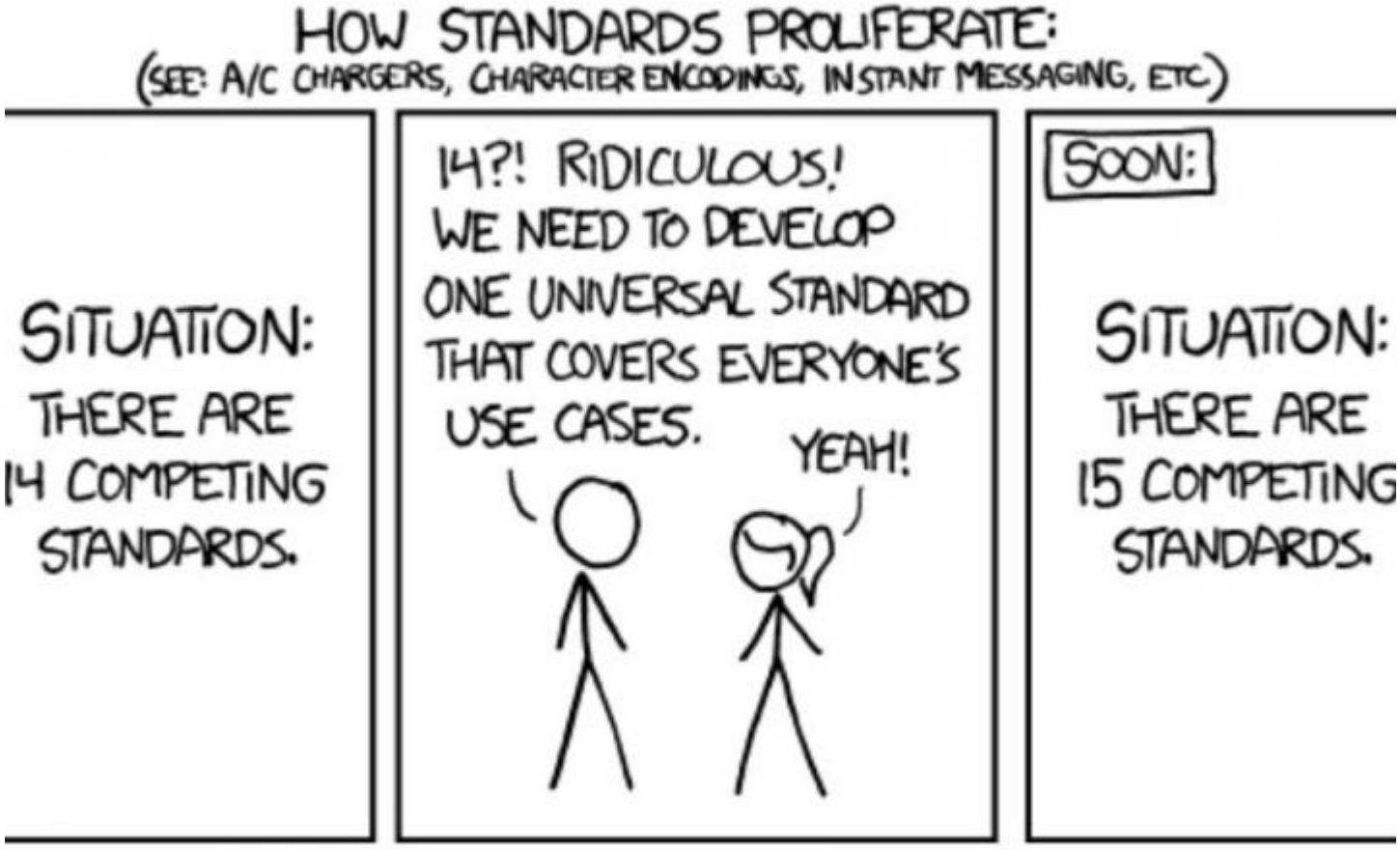
> ***DEFINITION*** - It is "a labeling, tagging or coding system used for recording cataloguing information or structuring descriptive records. A metadata schema establishes and defines data elements and the rules governing the use of data elements to describe a resource", Open data support

- One of the most commonly used metadata standard is Dublin Core.

Why standards?

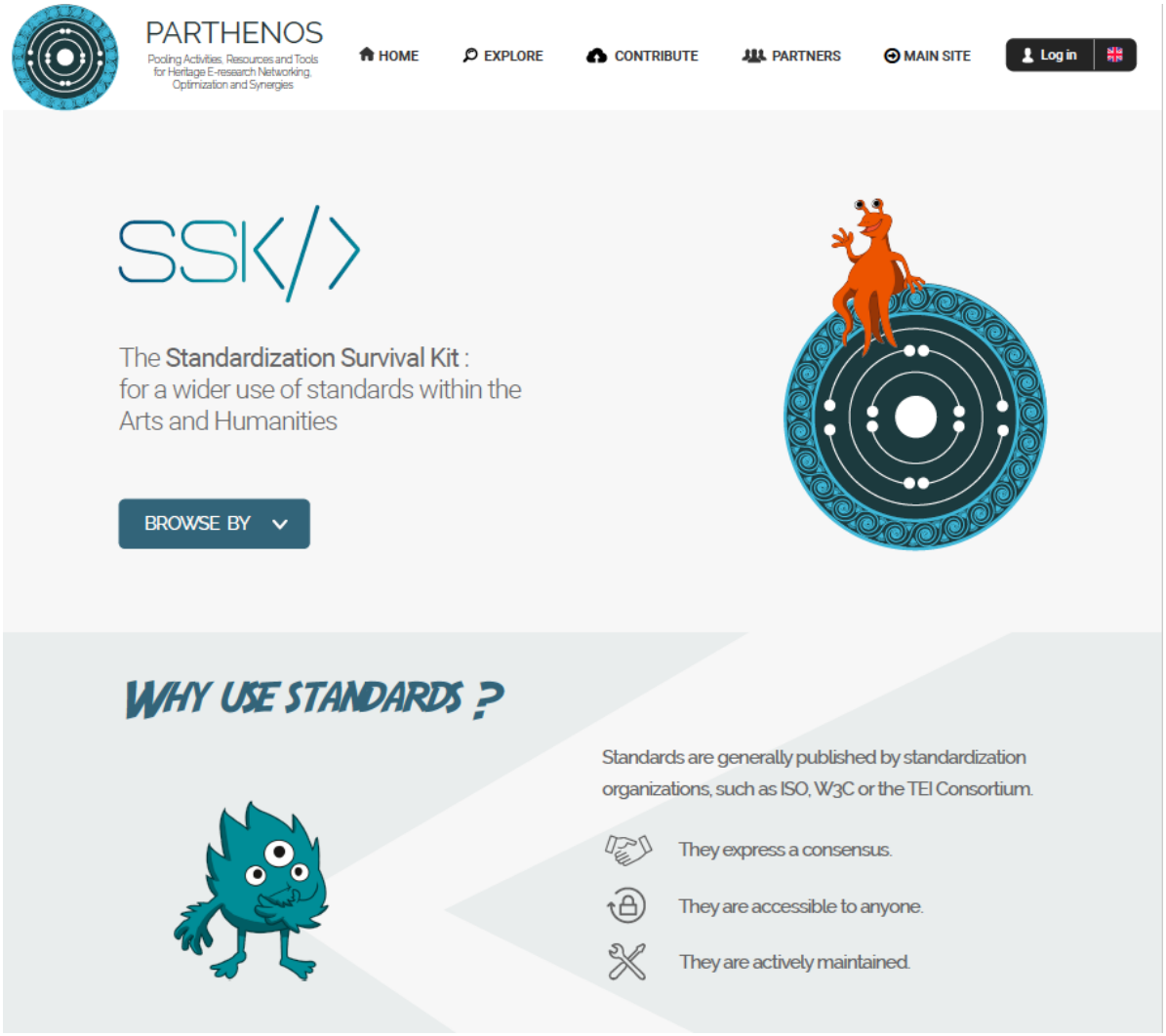


How to identify relevant standards for my research?



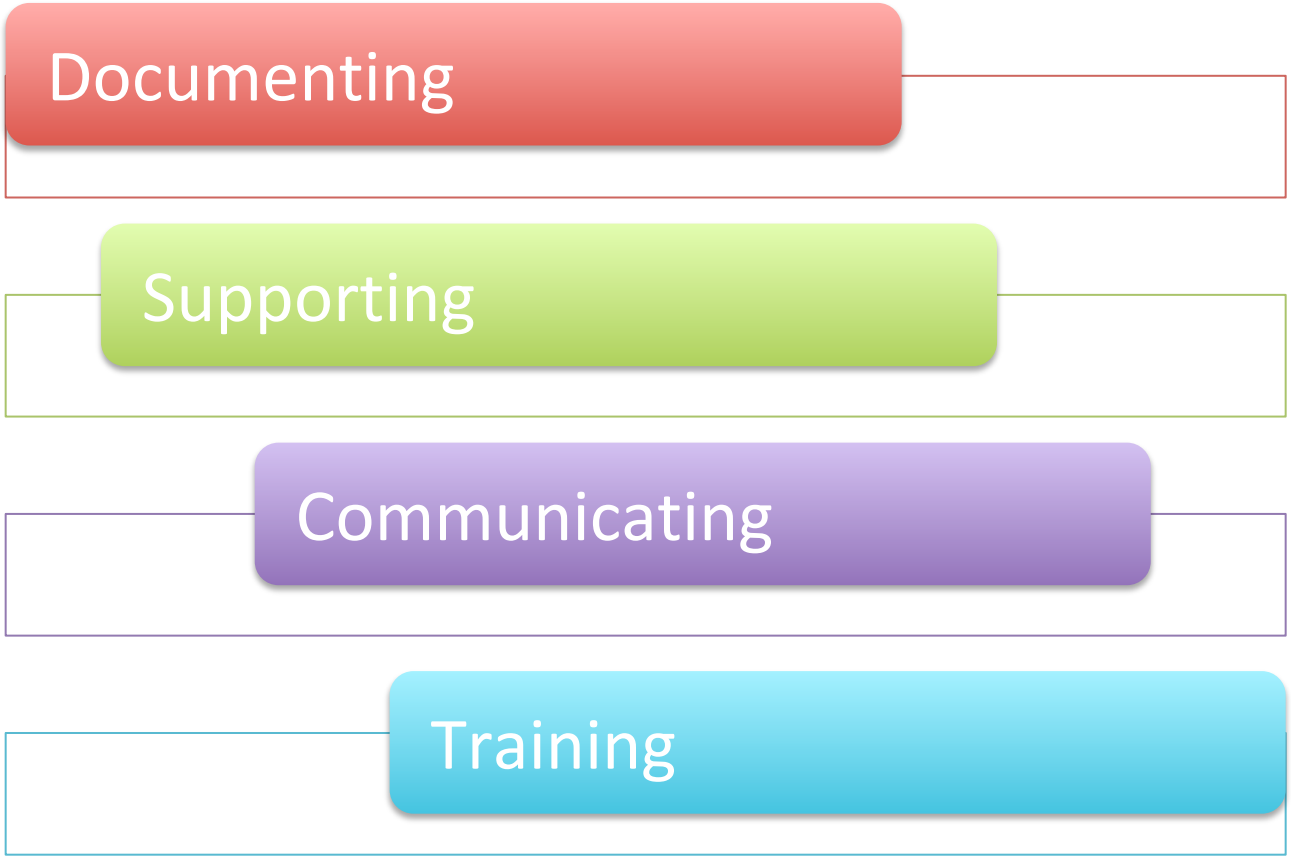
xkcd – "How standards proliferate"

The Standardization Survival Kit (SSK)



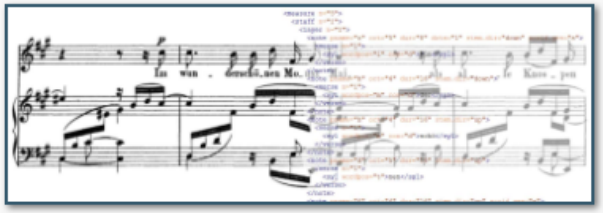
Homepage

The Standardization Survival Kit (SSK)



The Standardization Survival Kit (SSK)

Collaborative Digital Edition of a Musical Corpus



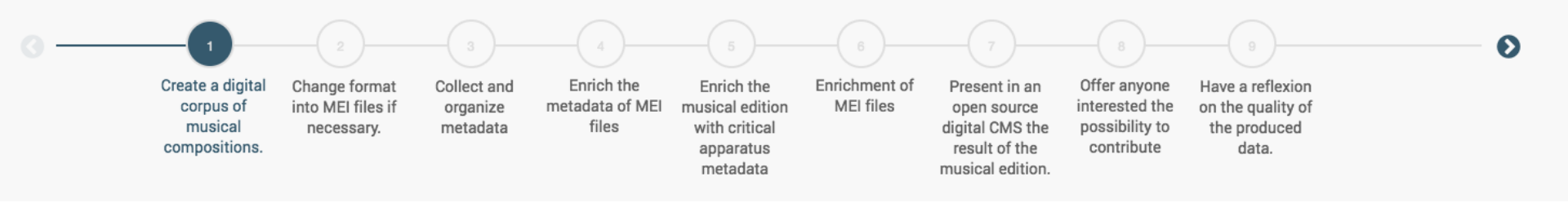
A project aims to do a digital edition of a musical corpus. The researchers need to be able to encode a broad range of musical documents in a machine-readable structure. The data to be encoded may include the musical content as provided by the composer (notes, pitches, durations, dynamics, etc.), information on the score (incipit, lyrics writer, etc.), information added by a performer when interpreting the content (timing, phrasing, various annotations, etc.), information on the visual appearance of the score (page layout, musical font, etc.) and analyses of the content in any of the other domains. The edition will be structured around a database in order to allow the users to explore it more easily. Furthermore, the project intends to be collaborative, which means it will offer anyone interested the possibility to contribute.

A research scenario (description)

PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

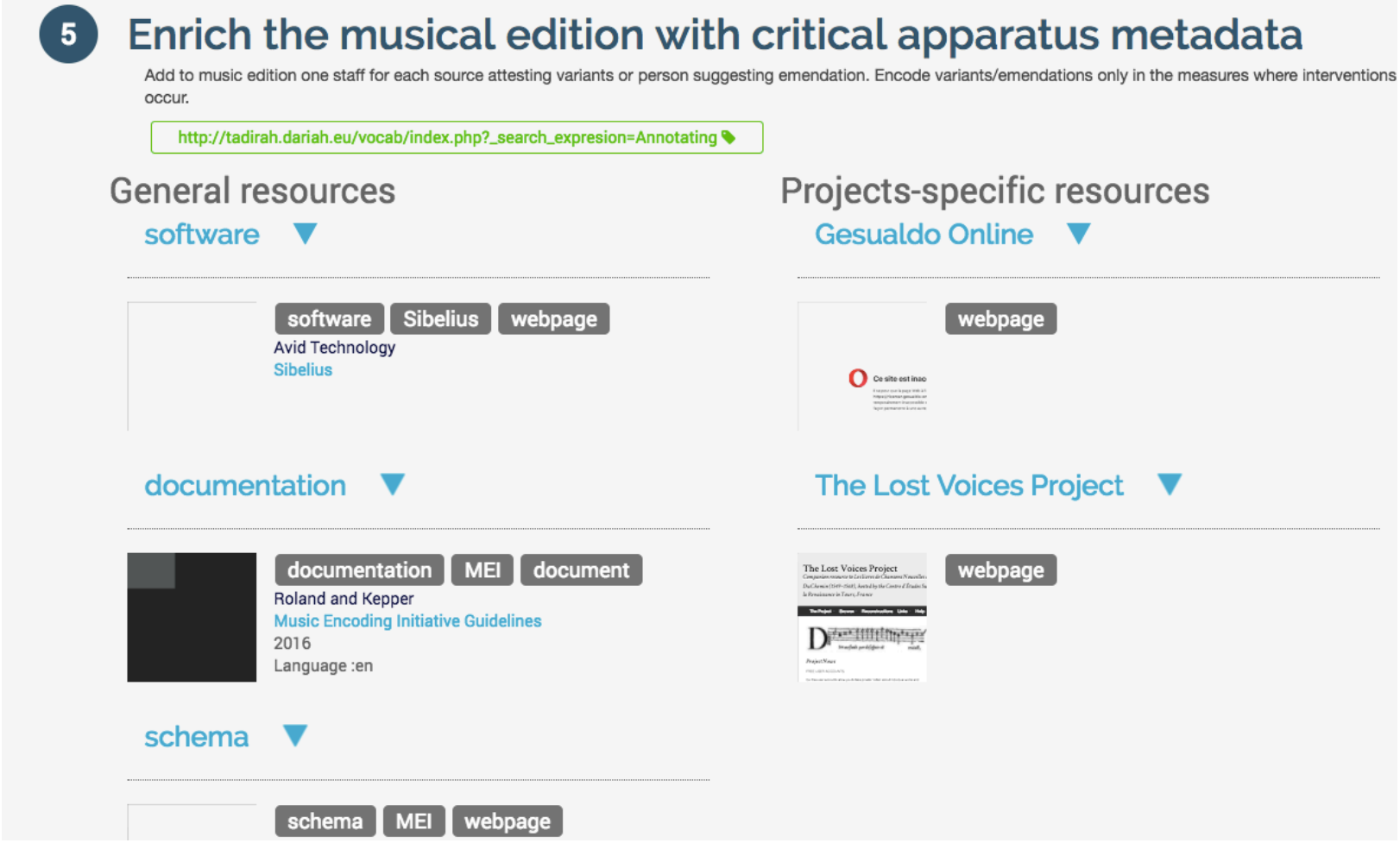
FH;P Fachhochschule Potsdam University of Applied Sciences

The Standardization Survival Kit (SSK)



A research scenario (steps)

The Standardization Survival Kit (SSK)



A research scenario (ressources)

The Devil is in the details!

- Improve your data and share them by using the services offered by **e-Infrastructures**
- Write a **DMP** to make your data FAIR
- Pay attention to "details", such as **standards**
- Choose the best suited standards for your data by using the **Standardization Survival Kit (SSK)**

Take away message

For better data,

don't mess with the Devil…

don't mess with **standards!**

Questions & Answers

(Picture CC0 https://pixabay.com/photo-2999583/)

Announcements

- **Next PARTHENOS webinar:**
  - "Research Infrastructures: Beyond tools – General introduction" (Steven Krauwer, Stefan Schmunk), 22.02.2018, 11:00 CET
- **New PARTHENOS Training Suite module:**
  Manage, Improve and Open Up Your Research and Data
- **New** clips with **Mork and Tork Cartoon**
  about Standards in several languages
- Launch **Standardization Survival Kit** (SSK) (soon!)
- **PARTHENOS Webinar Workshop** at the European Summer University in Digital Humanities 2018 (ESU), Leipzig

Questions & Answers

(Picture CC0 https://pixabay.com/photo-2999583/)

Thank you for joining us!

(Picture: CCO https://pixabay.com/photo-1889007/)

- ✓ Please help us improving the PARTHENOS Webinar Series and fill out the short Feedback Survey that we'll sent you in a follow-up e-mail.

PARTHENOS
Pooling Activities, Resources and Tools for Heritage E-research Networking, Optimization and Synergies

FH;P Fachhochschule Potsdam University of Applied Sciences

PARTHENOS eHumanities and eHeritage Webinar Series

Join the conversation on Twitter:
#PARTHENOSWebinar
@Parthenos_EU